Semester 1 Examinations 2023-2024

Course Instance Code(s)	4BCT, 1CSD1, 1CSD2, 1OA2, 4BS2, 1MAI1
Exam(s)	B.Sc. Computer Science and IT, M.Sc. in Data Analytics, M.Sc. in Artificial Intelligence, Science
Module Code(s)	CT4100
Module(s)	Information Retrieval
Paper No.	1
External Examiner(s)	Dr. R. Trestian
Internal Examiner(s)	Professor M. Madden
	*Dr. C. O'Riordan

Instructions: Answer any 3 questions. All questions are equally weighted.

Duration	2 hours
No. of Pages	3
Discipline(s)	Computer Science
Course Co-ordinator(s)	Dr. C. O'Riordan, Dr M. Nickles, Dr. F. Glavin

Requirements:

Release in Exam Venue	Yes
Handout	None
Statistical/ Log Tables	None
Cambridge Tables	None
Graph Paper	None
Log Graph Paper	None
Other Materials	None

Question 1

(a) With respect to text retrieval systems, outline, with examples, any three pre-processing steps often taken. Discuss the expected impact of these steps. (9)

(b) Describe briefly the vector space model for information retrieval. Discuss the strengths and limitations of this approach. (8)

(c) Outline a suitable approach to indexing a document collection to allow efficient handling of queries in a system adopting a vector space framework. (8)

Question 2

(a) Discuss an approach where user feedback (users identifying whether returned documents are relevant or not) can be used to extend or modify the query with the aim of improving performance. (7)

(b) Pseudo-feedback mechanisms have been used to augment queries or to suggest candidate terms to users. Suggest an approach to provide a diverse set of candidate terms to users to add to their queries. (8)

(c) Compare and contrast Rocchio's algorithm and pseudo-relevance feedback as methods for query expansion in information retrieval. Discuss the strengths, weaknesses, and situations where each method is more suitable. (10)

Question 3

(a) Precision and recall are often used to measure the quality of an answer set. These neglect some aspects of the user experience. Suggest a measure that may capture the user's experience. You may limit your answer to one aspect of the search process if you like. (8)

(b) Given a clustering algorithm, discuss a suitable approach to evaluating the quality of the algorithm. (8)

(c) Describe, with a suitable example, the main steps taken in collaborative filtering (recommender systems) to produce recommendations for a user. (9)

Question 4

(a) Many modern web-based search engines attempt to take into account the web link structure in addition to the content of the pages. Describe the Page Rank algorithm that uses information embedded in the web link structure to return relevant documents to a user. Discuss any limitations associated with this approach. (10)

(b) Explain three pre-retrieval measures that can be used to estimate the difficulty of a query. Include examples to illustrate your answer. (9)

(c) With reference to any information retrieval task of your choice, outline learning approaches that can be adopted to tackle these tasks. Outline the data available, the learning mechanism and any limitations of the approaches. (6)

END